

FOR THE RECORD

Similarity between pyridoxal/pyridoxamine phosphate-dependent enzymes involved in dideoxy and deoxyaminosugar biosynthesis and other pyridoxal phosphate enzymes



STEFANO PASCARELLA AND FRANCESCO BOSSA

Dipartimento di Scienze Biochimiche, A. Rossi Fanelli and Centro di Biologia Molecolare del C.N.R., Università La Sapienza, 00185 Rome, Italy

(RECEIVED December 10, 1993; ACCEPTED February 1, 1994)

Abstract: A multiple sequence alignment among aspartate aminotransferase, dialkylglycine decarboxylase, and serine hydroxymethyltransferase (DAS) was used for profile databank search. The DAS profile could detect similarities to other pyridoxal or pyridoxamine phosphate-dependent enzymes, like several gene products involved in dideoxsugar and deoxyaminosugar synthesis. The alignment among DAS and such gene products shows the conservation of aspartate 222 and lysine 258, which, in aspartate aminotransferase, interacts with the N1 of the coenzyme pyridine ring and forms the internal Schiff base, respectively. The lysine is replaced by histidine in the pyridoxamine phosphate-dependent gene products. The alignment indicates also that the region encompassing the coenzyme binding site is the most conserved.

Keywords: aminotransferase; antibiotics; deoxyaminosugars; dideoxsugars; profile analysis; pyridoxal phosphate; pyridoxamine phosphate; 3D structure superposition

The recent resolution of new spatial structures (Antson et al., 1993; Toney et al., 1993) indicates that the pyridoxal phosphate (PLP)-dependent enzymes involved in amino acid metabolism could be grouped in a few structural families, maybe only 2, exemplified by tryptophan synthase subunit β (Hyde et al., 1988) and aspartate aminotransferase (AAT; McPhalen et al., 1992); see Kinemage 1. However, within these 2 groups, sequence similarity among individual PLP-dependent enzymes is very low.

Accumulation and exploitation of structural and sequence information is enhancing the performance of databank search methods (e.g., profile analysis by Gribskov et al. [1990]; Vogt & Argos [1992]) for identifying distantly related proteins, as proven successfully in the case of PLP-dependent enzymes (Mehta et al., 1993, and references therein). These results prompted us

to compare the sequence of the *Escherichia coli* serine hydroxymethyltransferase (SHMT) to those of AAT and dialkylglycine decarboxylase (DGD; Toney et al., 1993) by combining a profile analysis (Genetics Computer Group, 1991) and a structural superposition approach (details in Pascarella et al., 1993). The resulting DGD-AAT-SHMT alignment (we will refer to it as DAS) proved to be sensitive in detecting distantly related PLP-dependent enzymes by profile search of the SWISS-PROT databank. In fact, the DAS profile search ranked the *eryC1* gene product (Dhillon et al., 1989; SWISS-PROT code ERBS_SACER) with a Z-score higher than that of several other PLP-dependent enzymes some of which definitely related to the AAT (Table 1), like the *Sulfolobus solfataricus* AAT (AAT_SULSO; fractional sequence identity to AAT is 18%). The alignment of ERBS_SACER to the DAS alignment obtained with PROFILEGAP is shown in Figure 1. Although sequence similarity is weak (pairwise identity percentage to AAT, DGD, and SHMT is 13, 14, and 12%, respectively), the lysine in position 258 (K258) of AAT, which forms a Schiff base with the PLP cofactor, and the aspartate in position 222 (D222 in AAT), which H bonds to the pyridine ring N1, are both conserved in the ERBS_SACER sequence (these residues will be referred to with their AAT numbering). These 2 residues are among those conserved in all aminotransferases (Mehta et al., 1993). The *eryC1* gene product is involved in the biosynthesis of erythromycin, very likely at the step of desosamine synthesis or its attachment to the macrolide ring (Dhillon et al., 1989). Previous studies suggested that amino sugar biosynthesis proceeds through a transamination step (Luckner, 1984). The similarity found with the DAS alignment supports the identification of ERBS_SACER with such an aminotransferase. The *eryC1* gene product has several homologues (Table 2). The PILEUP alignment among the first 7 sequences reported in Table 2 (indicated in the text as P7) showed that the aspartate of ERBS_SACER aligned to residue D222 of AAT is conserved in all of them, whereas residue K258 is replaced by histidine in only 1 case, i.e., in YRF7_SALTY. Two PLP-dependent enzymes were ranked

Reprint requests to: Francesco Bossa, Dipartimento di Scienze Biochimiche, Università La Sapienza, 00185 Rome, Italy; e-mail: pascarella@sci.uniroma1.it.

Table 1. PLP enzymes ranked at a score lower than that assigned to ERBS_SACER by profile search with DAS

| Target sequence ^a | Protein | Source | Z-score |
|------------------------------|--|------------------------------------|---------|
| ERBS_SACER | <i>eryC1</i> Gene product | <i>Saccharopolyspora erythraea</i> | 4.29 |
| AAT_SULSO | Aspartate aminotransferase | <i>Sulfolobus solfataricus</i> | 4.25 |
| BIOF_BACSH | 8-Amino-7-oxononoate synthase | <i>Bacillus sphaericus</i> | 4.21 |
| GSA_HORVU | Glutamate-1-semialdehyde 2,1-aminomutase | <i>Hordeum vulgare</i> | 4.19 |
| CEFD_STRCL | Isopenicillin N-epimerase | <i>Streptomyces clavuligerus</i> | 4.08 |

^a SWISS-PROT codes.

at the highest positions by the PROFILESEARCH routine run with the profile calculated from P7 (Table 3). For some of these gene products, a regulatory function was originally proposed on the basis of the effects of inactivation of the respective genes on metabolism. The predicted occurrence of a helix-turn-helix-like motif, typical of a group of expression-regulating factors, appeared to support this hypothesis (Lacalle et al., 1992; Stutzman-Engwall et al., 1992). On the contrary, our results suggest that the proteins mentioned in Table 2 are PLP-dependent enzymes.

More recently, several papers were published by Thorson and coworkers (Thorson et al., 1993, and references therein). These authors describe the cloning and characterization of a *Yersinia pseudotuberculosis* gene cluster involved in the biosynthesis of ascarylose, a 3,6-dideoxyhexose present in the O-antigen. One of these genes, *ascC*, codes for a pyridoxamine phosphate (PMP)-dependent iron-sulfur-containing enzyme (Gonzalez-Porque, 1986), which was in fact also purified and characterized. The same authors observed a sequence homology between the *ascC* product and the ERBS_SACER group members, which was

particularly high in the case of YRF7_SALTY (Thorson et al., 1993). It was proposed, on the basis of sequence similarity and putative reaction mechanism features, that these proteins were PLP-dependent aminotransferases involved in the biosynthesis of amino sugars. We included the 2 extra sequences reported by Thorson et al. (1993) (i.e., *ascC* and *tylB*, Table 2) in P7 (the resulting alignment will be indicated in the text as P9). In all, the aspartate D222 is conserved, whereas two of them, the PMP-dependent dehydrases *ascC* and YRF7_SALTY, have histidine instead of lysine K258. A profile search with P9 confirms and improves the results previously obtained with P7 (Table 3). Figure 1 reports the final sequence alignment among DAS and ERBS_SACER homologues. A dendrogram clustering ERBS_SACER sequences by sequence similarity shows clearly 3 groups (Fig. 2): one group collects the K258-containing putative aminotransferases, the second the PMP-dependent dehydrases containing the histidine, and the third coincides with YIFI_ECOLI, which has K258 but other peculiarities not yet well understood.

Table 2. Sequences homologous to ERBS_SACER

| Gene denomination ^a | Putative function | Reference |
|---|--|-------------------------------|
| <i>eryC1</i> (ERBS_SACER) | Erythromycin synthesis | Dhillon et al., 1989 |
| <i>degT</i> (DEGT_BACSU) | Pleiotropic regulatory gene | Takagi et al., 1990 |
| <i>dnrJ</i> (DNRJ_STRPE) | Daunorubicin synthesis | Stutzman-Engwall et al., 1992 |
| <i>prg1</i> | Puromycin synthesis | Lacalle et al., 1992 |
| <i>strS</i> | Streptomycin synthesis | Distler et al., 1992 |
| <i>orf10.4</i> in the <i>rffG-rffJ</i> intergenic region (YRF7_SALTY) | O antigen synthesis | Jiang et al., 1991 |
| <i>o299</i> in the <i>rffE-rffT</i> region (YIFI_ECOLI) | Enterobacterial common antigen synthesis | Daniels et al., 1992 |
| <i>ascC</i> | Ascarylose synthesis | Thorson et al., 1993 |
| <i>tylB</i> | Tylosin synthesis | Thorson et al., 1993 |

^a Available SWISS-PROT codes.

Fig. 1 (facing page). Sequence alignment obtained by extending the initial DAS-ERBS_SACER with iterated inclusion of DEGT_BACSU, DNRJ_STRPE, YRF7_SALTY, YIFI_ECOLI, and *prg1*, *strS*, *ascC*, *tylB* gene products. After each addition, a profile was recalculated from the resulting alignment and a new sequence incorporated with PROFILEGAP. AAT numbering system is reported in the top alignment line. Alignment has been manually modified to match the Ser/Thr residues in position 253 by adding the gap in position 254. AAT D222 and K258 are marked with “•”. The YIFI_ECOLI C-terminal part was unrealistically fragmented by PROFILEGAP and therefore the alignment corrected manually. Secondary structure assignments to AAT and DGD (line labeled with “ss”) according to Kabsch and Sander (1982) are reported: h, e, g, t, b, and s stand for helix, extended, β -turn, bulge, and bend, respectively. Blank means coil. Dots represent gaps introduced by the alignment routine. Absolute numbering for easy reference to Figure 3 is at the bottom alignment line.

Table 3. Highest scoring PLP-dependent enzymes detected with profile searches using probes P7 and P9

| Target sequence ^a | Protein | Source | Z-score with P7 | Z-score with P9 |
|------------------------------|---|----------------------------------|-----------------|-----------------|
| YDIM_HERAU | Hypothetical 34.7-kDa protein (homologous to cystathione lyase) | <i>Herpetosiphon aurantiacus</i> | 6.24 | 5.70 |
| GLYA_CAMJE | Serine hydroxymethyltransferase | <i>Campylobacter jejuni</i> | 4.86 | 4.93 |
| CGL_RAT | Cystathione- γ -lyase | <i>Rattus norvegicus</i> | 4.22 | 4.56 |
| KBL_ECOLI | 2-Amino-3-ketobutyrate coenzyme A ligase | <i>Escherichia coli</i> | 3.78 | 3.65 |
| DCLY_BACSU | Lysine decarboxylase | <i>Bacillus subtilis</i> | 3.59 | 3.60 |
| AAT_SULSO | Aspartate aminotransferase | <i>Sulfolobus solfataricus</i> | 3.58 | 4.36 |

^a SWISS-PROT codes.

The use of sensitive methods for sequence alignment and data bank searches with a multiple sequence alignment among distant proteins based on the 3-dimensional structure superposition has resulted in the identification and alignment of the ERBS_SACER proteins to PLP-dependent enzymes for which the spatial structures are known. The relatively high score given by the DAS profile to ERBS_SACER supports the hypothesis that these proteins also contain the "aminotransferase fold." This is confirmed by profile analysis with the probe P9 (ERBS_SACER plus homologues). In fact, it assigns to 1 PLP-dependent enzyme a Z-score >5.0, considered a threshold for definitive structural relationship, and to several others a score >3.0, con-

sidered meaningful (Mehta et al., 1993). The alignment among DAS and ERBS_SACER with its homologues indicates the presence of a lysine/histidine and an aspartate residue at positions equivalent to AAT K258 and D222, respectively (Fig. 1). The histidine at position 258 instead of the lysine which forms a Schiff base with the cofactor in the PLP- and PLP/PMP-dependent enzymes seems to occur only in the PMP-dependent enzymes. This is compatible with the fact that the latter forms only an "external" Schiff base between substrate and the PMP form of the cofactor, but they still need a base to catalyze a proton shift (Gonzalez-Porque, 1986). This central event in PLP- and PLP/PMP-dependent enzymes is supposed to be catalyzed by the ϵ -amino group of the same active site lysine residue (K258 in AAT, K87 in tryptophan synthase [Miles et al., 1989]). Interestingly, site-directed mutagenesis experiments on aspartate aminotransferases substituting the active site lysine with other polar amino acid residues, and in particular with histidine, showed that the mutated products maintain the proper fold and the capacity of binding PLP/PMP. Catalytic activity was still measurable but significantly lower than that of the native enzyme (Fukaki et al., 1990; Ziak et al., 1990). Similarly, experimental evidence obtained after mutation of the active site lysine in serine hydroxymethyltransferase suggests that this residue is not the base that removes a proton from the α carbon of glycine in its conversion to serine (Schirch et al., 1993). This confirms the hypothesis formulated by Christen that the ubiquitous active site lysine in PLP-dependent enzymes might "represent a consequence of the chemical reactivity of the cofactor rather than a mechanistic necessity" (Ziak et al., 1990). On the other hand, D222 is presumably required both by the PLP/PMP-dependent aminotransferases and the exclusively PLP- or PMP-dependent enzymes (such as SHMT and dideoxsugar dehydrases, respectively) because it interacts with the pyridine N1 and stabilizes the required positive charge during catalysis. An arginine equivalent to AAT arginine 386, which binds the α -carboxylate of the amino acid substrate, cannot be localized by sequence alignment in the other K258-bearing enzymes. Its presence would be expected in the enzymes involved in amino sugar biosynthesis because the amino group donor in the transamination reaction is still an amino acid. Very likely, the active site area of these enzymes was changed significantly during evolution in order to accommodate and bind sugar substrates. The similarity plot (Fig. 3) of the multiple sequence alignment in Figure 1 indicates that the region encompassing the putative PLP/PMP binding site (position 180–280 in Fig. 1) is most highly conserved. This region corresponds, according to the alignment in Figure 1, to

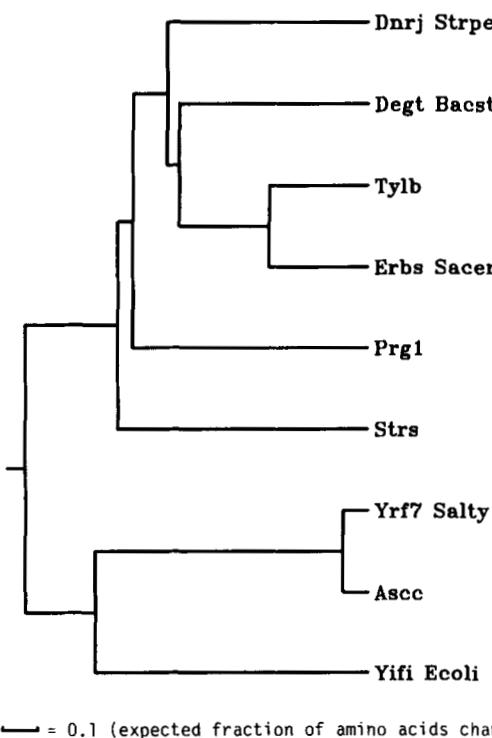


Fig. 2. Sequence similarity relationships among ERBS_SACER and its homologues. Sequence distances and the dendrogram have been calculated with the PHYLIP package (Felsenstein, 1985) using the programs PROTDIST and the UPGMA procedure implemented in NEIGHBOR, respectively. Branch lengths are proportional to sequence dissimilarity. Sequence identifiers are defined in the text.

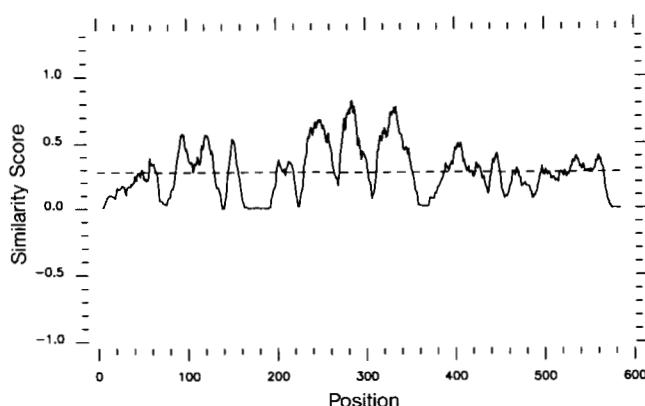


Fig. 3. PLOTSIMILARITY graph (Genetics Computer Group, 1991) showing the local similarity in the alignment in Figure 1 using a window of 10 residues. The most similar pairs, i.e., ERBS_SACER-*tylB* and *ascC-YRF7-SALTY*, were given a weight 0.5 (instead of 1.0 as for the other sequences) for correcting the overrepresentation of conserved strings. Average similarity is indicated by the dashed line and sequence position refers to the absolute numbering in Figure 1.

part of the AAT Rossmann-type α/β supersecondary structure (Kinemage 1; McPhalen et al., 1992). Included in the core of the large domain are β -strands *d*, *e*, *f*, *g* (AAT positions 184–188, 216–223, 250–255, and 268–273, respectively), and helices 8 and 9 (202–215, 233–246). Insertions/deletions in this region are also properly located in proximity of turns and loops, whereas in the N- and C-terminal part of the alignment they tend to interrupt or delete entire secondary structural modules. Alignment in such regions (which include the AAT small domain) is less reliable.

Relating the primary sequence of the dideoxy/deoxyamino-sugar biosynthetic enzymes to the known 3-dimensional structures of PLP enzymes and the availability of the corresponding genes provide a method to verify in future studies the structural and mechanistic similarity of these enzymes to the aspartate aminotransferase group by using site-directed mutagenesis.

Acknowledgments

Computer services provided by the Italian EMBnet node in Bari and by Progetto VAXRMA at University La Sapienza in Rome were essential. We thank Dr. Michael D. Toney and Prof. Johan H. Jansonius for providing the DGD atomic coordinates and Prof. V. Schirch for manuscript revision and helpful comments. This work was supported in part by a CNR grant in the framework of Progetto Speciale Bioinformatica and by a grant from the Commission of the European Communities (Human Capital and Mobility, contract ERBCHRXCT930179).

References

- Antson AA, Demidkina TV, Gollnick P, Dauter Z, Von Tersch RL, Long J, Berezhnoy SN, Phillips RS, Harutyunyan EH, Wilson KS. 1993. Three-dimensional structure of tyrosine phenol-lyase. *Biochemistry* 32:4195–4206.
- Daniels DL, Plunkett G III, Burland V, Blattner FR. 1992. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* 257:771–778.
- Dhillon N, Hale RS, Cortes J, Leadlay PF. 1989. Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. *Mol Microbiol* 3:1405–1414.
- Distler J, Mansouri K, Mayers G, Stockmann M, Piepersberg W. 1992. Streptomycin biosynthesis and its regulation in Streptomycetes. *Gene* 115: 105–111.
- Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- Fukaki S, Ueno H, Martinez del Pozo A, Pospisil MA, Manning JM, Ringe D, Stoddard B, Tanizawa K, Yoshimura T, Soda K. 1990. Substitution of glutamine for lysine at the pyridoxal phosphate binding site of bacterial D-amino acid transaminase. *J Biol Chem* 265:22306–22312.
- Genetics Computer Group. 1991. *Program manual for the GCG package, version 7*. Madison, Wisconsin.
- Gonzalez-Porque P. 1986. Cytidine diphosphate 4-keto-6-deoxy-D-glucose 3-dehydrogenase. In: Dolphin D, Poulton R, Avramovic O, eds. *Vitamin B₆ pyridoxal phosphate, chemical, biochemical and medical aspects, Part B*. New York: Wiley-Interscience. pp 391–419.
- Gribskov M, Lüthy R, Eisenberg D. 1990. Profile analysis. *Methods Enzymol* 183:146–159.
- Hyde CC, Ahmed SA, Padlan EA, Miles EW, Davies DR. 1988. Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*. *J Biol Chem* 263:17857–17871.
- Jiang XM, Neal B, Santiago F, Lee SJ, Romana LK, Reeves PR. 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar *typhimurium* (strain LT2). *Mol Microbiol* 5:695–713.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Lacalle RA, Tercero JA, Jiménez A. 1992. Cloning of the complete biosynthetic gene cluster for an aminonucleoside antibiotic, puromycin, and its regulated expression in heterologous hosts. *EMBO J* 11:785–792.
- Luckner M. 1984. *Secondary metabolism in microorganisms, plants, and animals*. Berlin: Springer-Verlag.
- McPhalen CA, Vincent MG, Jansonius JN. 1992. X-ray structure, refinement and comparison of three forms of mitochondrial aspartate aminotransferase. *J Mol Biol* 225:495–517.
- Mehta PK, Hale TI, Christen P. 1993. Aminotransferases: Demonstration of homology and division into evolutionary subgroups. *Eur J Biochem* 214:549–561.
- Miles EW, Kawasaki H, Ahmed SA, Morita H, Morita H, Nagata S. 1989. The β subunit of tryptophan synthase: Clarification of the roles of histidine 86, lysine 87, arginine 148, cysteine 170, and cysteine 230. *J Biol Chem* 264:6280–6287.
- Pascarella S, Schirch V, Bossa F. 1993. Similarity between serine hydroxymethyltransferase and other pyridoxal phosphate dependent enzymes. *FEBS Lett* 331:145–149.
- Schirch D, Delle Fratte S, Iurescia S, Angelaccio S, Contestabile R, Bossa F, Schirch V. 1993. Function of the active-site lysine in *Escherichia coli* serine hydroxymethyltransferase. *J Biol Chem* 268:23132–23138.
- Stutzman-Engwall KJ, Otten SL, Hutchinson CR. 1992. Regulation of secondary metabolism in *Streptomyces* spp. and overproduction of daunorubicin in *Streptomyces peucetius*. *J Bacteriol* 174:144–154.
- Takagi M, Takada H, Imanaka T. 1990. Nucleotide sequence and cloning in *Bacillus subtilis* of the *Bacillus stearothermophilus* pleiotropic regulatory gene *degT*. *J Bacteriol* 172:411–418.
- Thorson JS, Lo SF, Liu Hw. 1993. Biosynthesis of 3,6-dideoxyhexoses: New mechanistic reflections upon 2,6-dideoxy, 4,6-dideoxy, and amino sugar construction. *J Am Chem Soc* 115:6993–6994.
- Toney MD, Hohenester E, Cowan SW, Jansonius JN. 1993. Dialkylglycine decarboxylase structure: Bifunctional active site and alkali metal sites. *Science* 261:756–759.
- Vogt G, Argos P. 1992. Searching for distantly related protein sequences in large databases by parallel processing on a transputer machine. *Comput Appl Biosci* 8:49–55.
- Ziak M, Jaussi R, Gehring H, Christen P. 1990. Aspartate aminotransferase with the pyridoxal-5'-phosphate-binding lysine residue replaced by histidine retains partial catalytic competence. *Eur J Biochem* 187:329–333.